Name: _____

GCE Stats section 3

Population and samples

practice questions from GCE Maths

Date:

Question parts in red cover content in the GCE Statistics specification but outside this section of the scheme of work

Time:

Total marks available:

Total marks achieved: _____

Questions

Q1.

A company director decides to survey staff about changes to the company calendar. The company has staff in 4 different job roles

72 managers, 108 drivers, 180 administrators and 360 warehouse staff.

The director decides to take a stratified sample.

(a) Write down one advantage of using a stratified sample rather than a simple random sample for this survey.

(1)

(b) Find the number of staff in each job role that will be included in a stratified sample of 40 staff.

(3)

(c) Describe how to choose managers for the stratified sample.

(2)

(Total for question = 6 marks)

Q2.

(a) State two reasons why stratified sampling might be a more suitable sampling method than
simple random sampling.

(2)

(b) State two reasons why stratified sampling might be a more suitable sampling method than
quota sampling.

(2)

(Total for question = 4 marks)

Q3.

A mobile library has 160 books for children on its records. The librarian believes that books with fewer pages are borrowed more often. He takes a random sample of 10 books for children.

(a) Explain how the librarian should select this random sample.

(2)

The librarian ranked the 10 books according to how often they had been borrowed, with 1 for the book borrowed the most and 10 for the book borrowed the least. He also recorded the number of pages in each book. The results are in the table below.

Book	*A*	*B*	*C*	*D*	*E*	*F*	*G*	*H*	*I*	*J*
Borrowing rank	1	2	3	4	5	6	7	8	9	10
Number of pages	50	212	115	80	301	90	356	283	152	317

(b) Calculate Spearman's rank correlation coefficient for these data.

(4)

(c) Test the librarian's belief using a 5% level of significance. State your hypotheses clearly.

(3)

(Total 9 marks)

Q4.

(a) Explain what you understand by a random sample from a finite population.

(1)

(b) Give an example of a situation when it is not possible to take a random sample.

(1)

A college lecturer specialising in shoe design wants to change the way in which she organises practical work.

She decides to gather ideas from her 75 students.

She plans to give a questionnaire to a random sample of 8 of these students.

(c) (i) Describe the sampling frame that she should use.

(ii) Explain in detail how she should use a table of random numbers to obtain her sample.

(3)

(Total 5 marks)

Q5.

A gym club has 400 members of which 300 are males.

Explain clearly how a stratified sample of size 60 could be taken.

(3)

(Total 3 marks)

Q6.

A lake contains 3 species of fish. There are estimated to be 1400 trout, 600 bass and 450 pike in the lake. A survey of the health of the fish in the lake is carried out and a sample of 30 fish is chosen.

(a) Give a reason why stratified random sampling cannot be used.

(1)

(b) State an appropriate sampling method for the survey.

(1)

(c) Give one advantage and one disadvantage of this sampling method.

(2)

(d) Explain how this sampling method could be used to select the sample of 30 fish.

You must show your working.

(4)

(Total 8 marks)

Q7.

A nursery has 16 staff and 40 children on its records. In preparation for an outing the manager needs an estimate of the mean weight of the people on its records and decides to take a stratified sample of size 14.

(a) Describe how this stratified sample should be taken.

(3)

The weights, x kg, of each of the 14 people selected are summarised as

$$\sum x = 437 \text{ and } \sum x^2 = 26983$$

(b) Find unbiased estimates of the mean and the variance of the weights of all the people on the nursery's records.

(4)

(c) Estimate the standard error of the mean.

(2)

The estimates of the standard error of the mean for the staff and for the children are 5.11 and 1.10 respectively.

(d) Comment on these values with reference to your answer to part (c) and give a reason for any differences.

(2)

(Total for question = 11 marks)

Q8.

A college manager wants to survey students' opinions of enrichment activities. She decides to survey the students on the courses summarised in the table below.

Course	Number of students enrolled
Leisure and Sport	420
Information Technology	337
Health and Social Care	200
Media Studies	43

Each student takes only one course.

The manager has access to the college's information system that holds full details of each of the enrolled students including name, address, telephone number and their course of study. She wants to compare the opinions of students on each course and has a generous budget to pay for the cost of the survey.

(a) Give one advantage and one disadvantage of carrying out this survey using

(i) quota sampling,
(ii) stratified sampling.

(2)

The manager decides to take a stratified sample of 100 students.

(b) Calculate the number of students to be sampled from each course.

(3)

(c) Describe how to choose students for the stratified sample.

(2)

(Total 7 marks)

Q9.

A student believes that there is a difference in the mean lengths of English and French films. He goes to the university video library and randomly selects a sample of 120 English films and a sample of 70 French films. He notes the length, x minutes, of each of the films in his samples. His data are summarised in the table below.

	Σx	Σx^2	s^2	n
English films	10650	956909	98.5	120
French films	6510	615849	151	70

(a) Verify that the unbiased estimate of the variance, s^2, of the lengths of English films is 98.5 minutes^2

(2)

(b) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean lengths of English and French films are different.

(7)

(c) Explain the significance of the Central Limit Theorem to the test in part (b).

(1)

(d) The university video library contained 724 English films and 473 French films. Explain how the student could have taken a stratified sample of 190 of these films.

(3)

(Total 13 marks)

Q10.

A large company surveyed its staff to investigate the awareness of company policy. The company employs 6000 full time staff and 4000 part time staff.

(a) Describe how a stratified sample of 200 staff could be taken.

(3)

(b) Explain an advantage of using a stratified sample rather than a simple random sample.

(1)

A random sample of 80 full time staff and an independent random sample of 80 part time staff were given a test of policy awareness. The results are summarised in the table below.

	Mean score ($\bar{x}$)	Variance of scores (s^2)
Full time staff	52	21
Part time staff	50	19

(c) Stating your hypotheses clearly, test, at the 1% level of significance, whether or not the mean policy awareness scores for full time and part time staff are different.

(7)

(d) Explain the significance of the Central Limit Theorem to the test in part (c).

(2)

(e) State an assumption you have made in carrying out the test in part (c).

(1)

After all the staff had completed a training course the 80 full time staff and the 80 part time staff were given another test of policy awareness. The value of the test statistic z was 2.53

(f) Comment on the awareness of company policy for the full time and part time staff in light of this result. Use a 1% level of significance.

(2)

(g) Interpret your answers to part (c) and part (f).

(1)

(Total 17 marks)

Examiner's Report

Q1.

Part (a) made a reasonable start for many students. However, there were those who clearly understood the term 'stratified sampling', but missed the point of the question. The question required an advantage of stratified sampling. For example, some students stated that the proportions of each job role are the same in the sample as in the whole population. This explains how a stratified sample is created but not why.

Part (b) was answered very well by the majority of students.

There were many excellent answers to part (c) and typical answers in this category were succinct and precise. All the required detail was often contained in two short sentences. In contrast, there were some lengthy paragraphs from other students that did not always score both marks. A minority of students wrote vague responses. The question required some indication of how a random sample was to be achieved, using either random number tables or a random number generator on a calculator or a computer.

Q2.

For some students, the clarity of expression was poor, showing a lack of clear understanding and making it difficult to give credit. In part (a), the most common comments to gain credit related to representation of all groups, appropriateness for large samples or reflecting the population structure. The most common incorrect responses referred to large populations instead of samples or were too vague e.g. referring to better representation of 'the population' without reference to subgroups or structure. In part (b), common correct comments referred to interviewer bias for quota sampling and to calculation of sampling errors. Most common responses not scoring referred to sampling frames, recording of non-responses, or were too vague with simple reference to randomness

Q3.

A straightforward starter question dealing with correlation with most students managing the coefficient easily. They carried out the test well and typically gave its result in context.

The sampling only caused problems to those who are content with text book answers or who decided to go down the systematic route.

Q4.

This question was not answered particularly well, even by the best students. Problems and misunderstandings in describing how to take a sample were common and many students preferred the equal probability for each element approach in question (a). Marks were lost by students failing to be sufficiently precise in their descriptions in later parts. Many students gave an example rather than the lack of a sampling frame in question (b) .Only weaker students suggested systematic or quota sampling methods. In question (c)(i) many students failed to distinguish between the units and the list, whilst others were penalised for not specifying all the students. Generally the best answered part was question (c)(ii), however students need to pay attention to the detailed requirements.

Q5.

There were three stages required here and many candidates missed at least one of them. A

number of candidates failed to mention the need to label the males and females separately but most did state the need to select their samples using random numbers and the calculations (45 males and 15 females) were usually seen although a few only mentioned one group.

Q6.

This question proved to be a challenge to many candidates. There were a large number of responses that showed little understanding of how to carry out a survey or take a sample in practice. There were many answers that offered lists of advantages and disadvantages that related little, if at all, to the context of the question. The fact that a large number of candidates thought a practical solution involved enumerating the population and that many thought that they would 'interview' the fish showed how little this topic is understood.

Q7.

For all but the least able of students the calculations were easily and accurately carried out. However, very few students had any idea about what the comparison of standard errors was all about. Many of the students who scored poorly here would have used a standard error readily to calculate a sample size or confidence interval, but they have little or no idea of the underlying theory. The sampling was a source of easy marks to most apart from those who resorted to the text book again.

Q8.

Most candidates had a stab at part (a) but their responses were sometimes rather too brief or vague. Many quoted standard responses from textbooks or past mark schemes and, although a suitable pair of such answers was acceptable on this occasion candidate should really be aiming to address the question with reference to the context in the question. For example the access to a database means that a sampling frame is readily available.

The candidates were on much more familiar ground in part (b) and most scored full marks although a few failed to round 33.7 and 4.3 to suitable integers.

In part (c) there was some confusion with systematic sampling here and others failed to mention the need for a suitable sampling frame for each course although many did score the mark for using random numbers for the selection from each course.

Q9.

This question proved to be very well answered and it was pleasing to see so many fully correct solutions to question (a) and question (b), including conclusions in sufficient context to be awarded the marks available. question (c) and question (d), however, often lacked detail with values being calculated accurately in question (d) with no mention of a list being required. Students often stated 'at random' rather than describing how this might be achieved.

Q10.

Most candidates knew how to take a stratified sample by taking simple random samples in each stratum but they often forgot to describe how to label the members of the strata.

In (b) the commonest correct response was about the sample being more representative of the population but some missed the point and simply said that stratified sampling was "easier".

The calculation in part (c) was carried out very well by most candidates. There were few errors with the standard error and most correctly concluded that there was evidence of a difference in policy awareness between the types of staff.

In part (d) most knew that the Central Limit Theorem had something to do with the normal distribution but they did not mention that it was the mean scores of full time and part time staff that can be assumed to be normally distributed.

There were some correct responses to part (e) but many just mentioned independence despite this being given in the stem to part (c) of the question.

Most gave a correct conclusion in part (f) and some correctly inferred in the final part that the training course had been effective.

Some had the correct idea in part (g) although their conclusions went further than the evidence suggested: they claimed that the scores of the part time staff had increased, which may well be the case, but the evidence presented was only sufficient to conclude that the "gap" between policy awareness of the types of staff has been closed.

Mark Scheme

Section 3	Population and samples
qu.1	May 2017 S3 - Q1
qu.2	May 2016 S3 - Q1
qu.3	May 2015 S3 - Q1a
qu.4	May 2014 S3 - Q1
qu.5	June 2013 S3R - Q1
qu.6	June 2012 S3 - Q2
qu.7	May 2015 S3 - Q3a
qu.8	June 2013 S3 - Q3
qu.9	May 2014 S3R - Q5d
qu.10	June 2010 S3 - Q7ab

Q1.

Question Number	Scheme	Notes	Marks
(a)	-(accurate) estimates for each strata / job -**more** representative of the population -reflects population structure	Any 1 oe	B1
			(1)
(b)	Total staff=720	May be implied by calculations	B1
	Managers $= \frac{72}{720} \times 40 = 4$	For one correct calculation, follow through their 720.	M1
	Drivers $= \frac{108}{720} \times 40 = 6$		
	Administrators $= \frac{180}{720} \times 40 = 10$		
	Warehouse $= \frac{360}{720} \times 40 = 20$	4, 6, 10, 20 only. Must identify which job the values relate to.	A1
			(3)
(c)	Label all managers 1 – 72 o.e.	Idea of **sampling frame** or **list** of **managers**. Need not give the specific term.	B1
	Using random numbers in range 1-72 or 0-71 select 4 (managers).	Use of **random numbers** to select required number of managers. Must mention use of random numbers or some **random** selection process. If they are describing systematic sampling score B0.	B1
			(2)
			Total 6

Q2.

Question Number	Scheme	Marks
(a) e.g.	Analyse / find estimates for a particular **subgroup** of the population. Stratified guarantees **representation** of all **groups**, srs does not. Observe **relationships between subgroups** – srs does not guarantee equal or proportionate representation. Rare or extreme cases as part of a **small subgroups** can be represented **proportionately** in stratified i.e. stratified represents the structure of the population– srs does not allow this. Stratified typically require **large sample size** compared to srs due to lower variability within subgroups compared to entire population. Any 2 distinct reasons	B1B1 (2)
(b) e.g.	It (a stratified sample) is **not biased** as the members are chosen randomly. You can **estimate the sampling errors** (for a stratified sample) It (a stratrified sample) gives **more accurate estimates** as it is a random process. A quota sample may be (**interviewer / process**) **biased**. It's **not possible to estimate/find the sampling errors** for a quota sample (whereas you can for a stratified sample) Any 2 distinct reasons	B1B1 (2) **Total 4**
Notes	Award B1B1 two correct, B1B0 one correct. Allow 'it' for 'stratified'. Do not award marks for vague responses such as 'cheap', 'easy' 'quick' 'random' etc. Mentioning 'sampling frame' alone is not sufficient for a mark. Mentioning 'non-response are not recorded' alone is not sufficient for a mark.	

Q3.

Question Number	Scheme	Marks
(a)	Label all the books from 1 – 160 (o.e.) Use random numbers to select the 10 books	B1 B1 (2)
(b)	<table><tr><td>Book</td><td>A</td><td>B</td><td>C</td><td>D</td><td>E</td><td>F</td><td>G</td><td>H</td><td>I</td><td>J</td></tr><tr><td>Borrow rank</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td></tr><tr><td>Page rank</td><td>1</td><td>6</td><td>4</td><td>2</td><td>8</td><td>3</td><td>10</td><td>7</td><td>5</td><td>9</td></tr><tr><td>d^2</td><td>0</td><td>16</td><td>1</td><td>4</td><td>9</td><td>9</td><td>9</td><td>1</td><td>16</td><td>1</td></tr></table> $r_s = 1 - \frac{6\times 66}{10(100-1)}$ $[= 1 - 0.4]$ $= 0.6$ <u>0.6</u>	M1 M1 M1,A1 (4)
(c)	H_0: $\rho = 0$ H_1: $\rho > 0$ Critical value is 0.5636 0.6 > cv so significant result and sufficient evidence to reject H_0 There is support for the librarian's belief <u>or</u> there is evidence of a correlation between the number of pages in a book and the number of times it is borrowed.	B1 B1 B1ft (3) **Total 9**

Notes

(a)

1st B1 for labelling\numbering\listing\using sampling frame of all 160 books

2nd B1 for use of random numbers\selection and mentioning the number 10

(b)

1st M1 for an attempt to rank the number of pages (at least 4 correct) Allow reverse ranks

2nd M1 for attempt at d^2 row (may be implied by sight of $\sum d^2 = 66$ or 264 for reverse ranks)

3rd M1 for use of the correct formula, follow through their $\sum d^2$ if clearly stated

If answer is not correct, a correct expression is required.

A1 for 0.6 (or -0.6 for reverse ranks)

(c)

1st B1 for both hypotheses in terms of ρ, one tail H_1 (compatible with ranks) Allow use of ρ_s

Hypotheses just in words e.g. "no correlation" score B0.

2nd B1 for cv of 0.5636 [If they have a two tail H_1 then allow 0.6485]

Allow $\pm$ for reverse ranking but must be same sign as r_s

If hypotheses are the wrong way around this must be B0 but 3rd B1 is possible.

3rd B1ft for a correct contextualised comment. Must mention "librarian" (or he)

<u>or</u> "number of pages" and "borrowing"

Follow through their r_s and their cv (provided it is |cv| <1)

Don't insist on the word "positive" or "negative" for a one-tailed test

Use of "association" is B0

Independent of 1st B1 so if $|r_s| >$ |cv| must say there is sufficient evidence of …….(o.e.)

and if $|r_s| <$ |cv| must say insufficient evidence of … (o.e.) regardless of their hypotheses

Q4.

Question Number	Scheme	Marks
(a)	(This is a sample where) **every** (possible) **sample** (of size n) has an **equal chance** of being chosen.	B1 (1)
(b)	'When it is impossible to provide a **sampling frame**' or a correct example with an indication of sampling frame being impossible.	B1 (1)
(c)(i)	A **list/register of all** the students.	B1
(ii)	Number the students (from 0 to 74, 1 to 75 etc.) Using the **random no. table** read off the nos. and **identify or select the students allocated those nos.**	B1 B1 (3) **Total 5**
	Notes	
(a)	Require **all / each / every** etc **sample** and **same/equal** etc **chance / probability** etc for B1	
(b)	Require **impossible / no / doesn't exist** etc and **sampling frame** for B1	
(c)(i)	Require **list/register etc** and **all/every/75 etc students** for B1 List of 8 students is B0	
(ii)	First B1 accept **'in the corresponding position' o.e.** if numbering omitted Second B1 require both for mark.	

Q5.

Question Number	Scheme	Marks
	Label females 1 – 100 (or 0 – 99) and males 1 – 300 (or 0 – 299) Using *random numbers* for each group in range 1 – 100 (0 – 99) select 15 females and using 1 – 300 (or 0 – 299) select 45 males	B1 B1 B1 **[Total 3]**
	Notes	
	1st B1 for labelling\numbering\listing females and males 2nd B1 for use of random numbers or "randomly select" in each group (may be implied) 3rd B1 for selecting the correct number of females and males e.g. randomly select 45 males and 15 females scores 2nd and 3rd B marks since randomly selecting and the "each group" is implied, If using systematic sampling within each strata allow 1st B1 and 3rd B1 (if earned) but 2nd B0	

Q6.

Question Number	Scheme	Marks
(a)	Sampling frame within each species of fish in the lake impossible to obtain.	B1 **(1)**
(b)	Quota sampling	B1 **(1)**
(c)	Advantages: Sample can be obtained quickly Costs are kept to a minimum Administration of survey is easy Disadvantages: Not possible to estimate sampling errors Process not random Surveyor may not be able to identify species of fish easily	B1 B1 **(2)**
(d)	<table><tr><th>Species</th><th>Quota</th></tr><tr><td>Trout</td><td>$\frac{1400}{2450}\times 30 = 17.14$</td></tr><tr><td>Bass</td><td>$\frac{600}{2450}\times 30 = 7.35$</td></tr><tr><td>Pike</td><td>$\frac{450}{2450}\times 30 = 5.51$</td></tr></table> Fish are caught from the lake until the quota of 17 trout, 7 bass and 6 pike are reached. If a fish is caught and the species quota is full, then this is ignored.	B1B1B1 B1 (4) **Total 8**

Notes

(a) 'You can't / it's very difficult to number all the fish' or equivalent

(c) Correct answer to (b) required. Some detail required.

(d) 1st B1 any one correct calculation seen or implied
2nd B1 all correct to at least 1 dp
3rd B1 for 17,7,6
4th B1 accept equivalent statement. Require comment on what to do with 'extra fish'.

Q7.

Question Number	Scheme	Marks
(a)	Label staff (from 1 – 16) and children (from 1 – 40) Use random numbers to select 4 staff and 10 children	B1 B1 B1 (3)
(b)	$\bar{x} = \hat{\mu} = 31.2142\ldots$ **awrt** **<u>31.2</u>** $s^2 = \frac{26983 - 14 \times "31.2..."^2}{13}$ $= 1026.33\ldots$ **awrt** **<u>1030</u>**	B1 M1 A1ft A1 (4)
(c)	$\frac{"\sqrt{1026.33...}"}{\sqrt{14}}$, $= 8.562..$ **awrt** **<u>8.56</u>**	M1, A1 (2)
(d)	The variation within each stratum is quite small (o.e.) **The difference in the means will be quite large**, (so variations from the overall mean will be large giving a larger overall s.e.)	B1 B1 (2) **Total 11**

Notes

(a)

1st B1 for labelling\numbering\listing staff <u>and</u> children

2nd B1 for use of random numbers or "randomly select" in <u>each group</u> (may be implied)

3rd B1 for selecting the correct number of staff <u>and</u> children

e.g. randomly select 4 staff and 10 children scores 2nd and 3rd B marks since randomly selecting and the "each group" is implied,

(b)

B1 for awrt 31.2

M1 for a correct expression ft their $\bar{x}$ and allow transcription error in $\sum x^2$ e.g. 29683

1st A1ft for a fully correct expression ft their $\bar{x}$ only

2nd A1 for awrt 1030

(c)

M1 for attempting $\frac{"\text{their } s"}{\sqrt{14}}$ (must have 14)

A1 for awrt 8.56

(d)

1st B1 for a suitable comment about **variation** (se) suggesting that **variation** (se) **within strata is less than that overall**

2nd B1 for a suitable **reason** about **means**, pointing out that the individuals' weights will vary a lot from the overall mean and so overall s.e. will be higher.

Q8.

Question Number	Scheme	Marks
(a)i e.g e.g	**Quota Sampling:** Advantages: Fieldwork can be done **quickly**, or administering the test is **easy**, or costs are kept to a minimum (**cheap**), or gives estimates for each course. or OK for large populations or sampling frame not required (o.e.) Disadvantages: **Non-random** process or not possible to estimate the sampling errors, or non response not recorded, or interviewer can introduce **bias** in sample choice. (o.e.)	B1
(a)ii e.g. e.g.	**Stratified Sampling:** Advantages: Can give accurate estimates as it is a **random** process, or gives estimates for each course or **representative** of [BUT not "proportional" to] the whole population. (o.e.) Disadvantages: Sampling frame required, or strata may not be clear as some students overlap courses or not suitable for large populations. (o.e.)	B1 (2)
(b)	Total enrolments=1000 (may be implied by calculations) Leisure and Sport=$\frac{420}{1000}\times 100=42$ Information Technology=$\frac{337}{1000}\times 100=33.7=34$ Health and Social Care=$\frac{200}{1000}\times 100=20$ Media Studies=$\frac{43}{1000}\times 100=4.3=4$	B1 M1 A1 (3)
(c)	The college's information system would be used to identify each student and which course they are enrolled on. i.e. idea of **sampling frame** or **list** for **each course**. Use of **random numbers** to select required number of students **from each course**	B1 B1 (2) **Total 7**

	Notes
	Do not penalise for lack of context in part (a)
(a)	1st B1 for an advantage and a disadvantage for quota sampling (must be 1st or labelled (i)) 2nd B1 for an advantage and a disadvantage for stratified sampling (2nd or labelled (ii)) Do not allow opposite pairs e.g. "quicker/easier" for quota sampling and "takes a long time/more difficult" for stratified or quota "easy to use" but strat. "hard for large populations" Do not allow same reason for both e.g. "gives estimates for each course"
(b)	M1 for one correct calculation, ft their "1000" A1 for 42, 34, 20 and 4 only
(c)	1st B1 for some mention of a suitable sampling frame. Need not give the specific term but a suitable source of list is required for all students in each course. 2nd B1 for mentioning use of random numbers or some random selection process for each course. If they are describing systematic sampling score B0 here

Q9.

Question Number	Scheme	Marks
(a)	$S_E{}^2 = \frac{1}{n-1}\left(\sum x^2 - \frac{(\sum x)^2}{n}\right) = \frac{1}{119}\left(956909 - \frac{10650^2}{120}\right)$	M1
	$= \frac{11721.5}{119} = 98.5$	A1
		(2)
(b)	H_0: $\mu_F = \mu_E$, H_1: $\mu_F \neq \mu_E$,	B1
	$\bar{x}_E = \frac{10650}{120} = 88.75$ and $\bar{x}_F = \frac{6510}{70} = 93$	M1
	Test statistic, $z = \frac{93-88.75-0}{\sqrt{\frac{151}{70}+\frac{98.5}{120}}} = 2.4627\ldots$	M1A1
	Critical values, $z = (\pm)2.5758$	B1ft
	Test stat is not in critical region	M1
	Insufficient evidence to reject H_0 at 1% level	A1ft
	No significant evidence of a difference in mean lengths of English and French films	(7)
(c)	By CLT we can assume that the mean of a large sample has a Normal distribution	B1 (1)
(d)	On a list, label English films 1 – 724 and French films 1-473 (oe)	B1
	Use random number table/generator to select	
	$\frac{724}{724+473} \times 190 = 115$ English films and	M1A1
	$\frac{473}{1197} \times 190 = 75$ French films	(3)
		(13 marks)

Notes		
	Alternative	
(a)	$S_E{}^2 = \frac{n}{n-1}\left(\frac{\sum x^2}{n} - \bar{x}^2\right) = \frac{120}{119}\left(\frac{956909}{120} - 88.75^2\right) = 98.5$	
(b)	1st B1 needs both H_0 and H_1, can be in words 2nd B1ft on their H_1 1st M1 for attempt @ both means ($\bar{x}_E$ may be in (a)) 2nd M1 for attempt at correct test statistic, ft their values 3rd M1 for attempt to compare their test stat and critical values A1 ft on their test and critical values but must include comment in context	
(c)	Require mention of mean of *E* or *F* and normal distribution	
(d)	M1 requires use of <u>random numbers</u> and attempt to find correct sample sizes A1 both 115 and 75 found.	

Q10.

Question Number	Scheme	Marks
(a)	Label full time staff 1-6000, part time staff 1-4000	M1
	Use random numbers to select	M1
	Simple random sample of 120 full time staff and 80 part time staff	A1 (3)
(b)	Enables estimation of statistics / errors for each strata <u>or</u> "reduce variability" <u>or</u> "more representative" <u>or</u> "reflects population structure" **NOT** "more accurate"	B1 (1)
(c)	$H_0: \mu_f = \mu_p$, $H_1: \mu_f \neq \mu_p$ (accept μ_1, μ_2)	B1
	s.e. $= \sqrt{\frac{21}{80}+\frac{19}{80}}$, $z = \frac{52-50}{\sqrt{\frac{21}{80}+\frac{19}{80}}} = \left(2\sqrt{2}\right)$	M1,M1
	$= 2.828\ldots$ (awrt **2.83**)	A1
	Two tailed critical value z = 2.5758 (or prob of awrt 0.002 (<0.005) or 0.004 (<0.01))	B1
	[2.828 > 2.5758 so] significant evidence to reject H_0	dM1
	There is evidence of a difference in policy awareness between full time and part time staff	A1ft (7)
(d)	Can use mean full time and mean part time	B1
	~ Normal	B1 (2)
(e)	Have assumed $s^2 = \sigma^2$ or variance of sample = variance of population	B1 (1)
(f)	2.53 < 2.5758, not significant <u>or</u> do not reject H_0	M1
	So there is insufficient evidence of a difference in mean awareness	A1ft (2)
(g)	Training course has closed the gap between full time staff and part time staff's mean awareness of company policy.	B1 (1)
		17

(a)	1st M1 for attempt at labelling full-time and part-time staff. One set of correct numbers. 2nd M1 for mentioning use of random numbers 1st A1 for s.r.s. of 120 full-time and 80 part-time
(c)	1st M1 for attempt at s.e. - condone one number wrong . NB correct s.e. = $\sqrt{\frac{1}{2}}$ 2nd M1 for using their s.e. in correct formula for test statistic. Must be $\frac{\pm(52-50)}{\sqrt{\frac{p}{q}+\frac{r}{s}}}$ 3rd dM1 **dep. on 2nd M1** for a correct statement based on their normal cv and their test statistic 2nd A1 for correct comment in context. Must mention "scores" or " policy awareness" and types of "staff". Award **A0** for a one-tailed comment. Allow ft
(d)	1st B1 for mention of mean(s) <u>or</u> use of $\bar{X}$, provided $\bar{X}$ clearly refers to full-time or part-time 2nd B1 for stating that distribution can be assumed normal e.g. "mean score of the test is normally distributed" gets B1B1
(f)	M1 for correct statement (may be implied by correct contextualised comment) A1 for correct contextualised comment. Accept "no difference in mean scores". Allow ft
(g)	B1 for correct comment in context that implies training was effective. This must be supported by their (c) and (f). Condone one-tailed comment here.